



Do High-Stakes Placement Exams Predict College Success?

Judith Scott-Clayton

February 2012

CCRC Working Paper No. 41

Address correspondence to:

Judith Scott-Clayton
Assistant Professor of Economics and Education and
Senior Research Associate, Community College Research Center
Teachers College, Columbia University
525 West 120th Street, Box 174
New York, NY 10027
212-678-3091
Email: scott-clayton@tc.columbia.edu

This project was funded by the Bill & Melinda Gates Foundation. Olga Rodriguez and Michelle Hodara provided expert assistance in cleaning, coding, and interpreting the data. The analyses presented, interpretations made, and views expressed are solely the responsibility of the author and do not necessarily represent the views of the funding or data-providing agencies. All errors are my own.

Abstract

Community colleges are typically assumed to be nonselective, open-access institutions. Yet access to college-level courses at such institutions is far from guaranteed: the vast majority of two-year institutions administer high-stakes exams to entering students that determine their placement into either college-level or remedial education. Despite the stakes involved, there has been relatively little research investigating whether such exams are valid for their intended purpose, or whether other measures of preparedness might be equally or even more effective. This paper contributes to the literature by analyzing the predictive validity of one of the most commonly used assessments, using data on over 42,000 first-time entrants to a large, urban community college system. Using both traditional correlation coefficients as well as more useful decision-theoretic measures of placement accuracy and error rates, I find that placement exams are more predictive of success in math than in English, and more predictive of who is likely to do well in college-level coursework than of who is likely to fail. Utilizing multiple measures to make placement decisions could reduce severe misplacements by about 15 percent without changing the remediation rate, or could reduce the remediation rate by 8 to 12 percentage points while maintaining or increasing success rates in college-level courses. Implications and limitations are discussed.

Contents

1. Introduction.....	1
2. National Testing Context.....	2
3. Theoretical Background and Previous Research.....	4
3.1 Concepts of Test Validity.....	4
3.2 Evidence Regarding Predictive Validity	5
4. The Institutional Context and Data	8
4.1 Institutional Context.....	8
4.2 Data and Sample	9
5. Empirical Strategy and Main Results	12
5.1 Predictor Variables and Success Criteria	12
5.2 Analysis of Variation	14
5.3 Predicted Placement Accuracy Rates and Incremental Accuracy Rates of Placement Tests.....	19
5.4 All Mistakes Are Not Equal: Minimizing the Severe Error Rate and Other Considerations.....	24
5.5 Comparing Placement Outcomes Across Alternative Sets of Predictors	29
5.6 Summary of Empirical Results	32
6.1 Possible Explanations for the Limited Predictive Validity of Placement Exams ..	33
6.2 Limitations of the Use of High School Background Measures.....	35
6.3 The Salience of Different Types of Placement Mistakes.....	35
6.4 The Impact of Remediation on Future College-Level Outcomes	36
6.5 Summary and Conclusions.....	37
References.....	39

1. Introduction

Community colleges are typically assumed to be nonselective, open-access institutions, yet access to college-level courses at such institutions is far from guaranteed. Instead, many students' first stop on campus will be to an assessment center where they will take exams in math, reading, and/or writing. The vast majority (92 percent) of two-year institutions administer these high-stakes exams to help determine who may enroll in college-level courses and who will be referred to remedial education (Parsad, Lewis, & Greene, 2003).¹ Often, placement is determined solely on the basis of whether a score is above or below a certain cutoff.

For the majority of students at community colleges, the consequence of assessment is placement into remediation in at least one subject. A recent study of over 250,000 students at 57 community colleges across the country found that 59 percent were referred to developmental math and 33 percent were referred to developmental English (Bailey, Jeong, & Cho, 2010). Students must pay tuition for remedial courses, but the credits they earn do not count toward graduation requirements. The cost to schools of providing this remedial instruction has been estimated at \$1 billion or more (Noble, Schiel, & Sawyer, 2004).

Unfortunately, the remedial "treatment" that is assigned on the basis of these assessments is not obviously improving outcomes. Bailey et al. (2010) found that students who ignored a remedial placement and instead enrolled directly in a college-level class had slightly lower success rates than those who placed directly into college-level, but substantially *higher* success rates than those who complied with their remedial placement, because relatively few students who entered remediation ever even attempted the college-level course.² In addition, of several studies using quasi-experimental designs to estimate the impact of remediation, only one indicates positive effects while three others have found mixed or even negative results (Bettinger & Long, 2009; Calcagno & Long, 2008; Martorell & McFarlin, 2011; Boatman & Long, 2010). This raises questions not only about the effectiveness of remedial instruction, but also about the entire process by which students are assigned to remediation.

¹ Throughout, I use the terms "remedial" and "developmental" interchangeably.

² "Success" here is defined as passing the first college-level class, or "gatekeeper" class.

Despite the stakes involved, the validity of these exams has received relatively little attention. A Google search for “+validity ACT SAT” returns 2.8 million results, while an equivalent search for the two most commonly used placement exams, the COMPASS (published by ACT, Inc.) and the ACCUPLACER (by the College Board), returns just 4,610 results. And while there is a long history of empirical research into the predictive validity of college entrance exams, only a handful of studies have examined these high-stakes college placement exams. Most of these studies have been conducted by the test makers themselves.

This paper contributes to the literature by analyzing the predictive validity of one of the most commonly used assessments, using data on over 42,000 first-time entrants to a Large Urban Community College System (LUCCS).³ I analyze both standard statistical measures of predictive power (such as correlation coefficients) as well as more tangible decision-theoretic measures that may be more useful for policy decisions, including absolute and incremental placement accuracy rates (that is, the percent of students predicted to be accurately placed under a given set of tests and rules) and a new measure I call a *severe error rate*. Importantly, I examine whether other measures of preparedness, such as high school background, might be equally or even more predictive of college success.

The following section describes the testing context nationally. Section 3 describes the theoretical background and previous literature relating to placement test validity. Section 4 describes the institutional context and data. Section 5 presents the empirical strategy and main results. Section 6 presents extensions and robustness checks, and Section 7 concludes with a discussion of potential policy implications.

2. National Testing Context

Nationally, two college placement exams dominate the market: ACCUPLACER®, developed by the College Board, is used at 62 percent of community colleges, and COMPASS®, developed by ACT, Inc., is used at 46 percent (Primary Research Group, 2008). These percentages are not mutually exclusive, as some schools

³ The system requested anonymity.

may “mix and match” depending on the test subject. Both testing suites include a written essay exam, an ESL exam, and computer-adaptive tests in reading comprehension, writing/sentence skills, and several modules of math from arithmetic to trigonometry. Schools can choose from these exams “à la carte.” While these are the most commonly used tests, several states, including Texas and Florida, have also worked with testing companies to develop their own exams. As will be described below, LUCCS uses several standard COMPASS exams as well as a customized writing exam developed in partnership with ACT, Inc.

Because most of the test modules are adaptive (meaning that questions are tailored to different test takers depending on their responses to previous questions), these tests may be very short. For example, scores on a COMPASS algebra exam may be determined by as few as eight questions (ACT, Inc., 2006, p. 91). The tests are not timed, but on average each test component takes about 30 minutes to complete, such that an entire suite of placement exams may be completed in two hours or less (College Board, 2007; ACT, Inc., 2006).⁴

Although recent years have seen a trend toward increasing standardization in how placement exams are used, practices still vary greatly from state to state, system to system, and school to school (see a recent review by Hughes & Scott-Clayton, 2011). Tests may be mandatory upon entry, or students may be allowed to defer them and still take some introductory-level courses in the meantime. Students may be exempted based upon ACT/SAT scores, high school test scores, or field of study (for example, some career-technical programs may not require testing, or may use an entirely different test). Placement decisions may be based solely on test scores, may incorporate additional information, or may be entirely at the discretion of the student. The cutoff scores that determine placement often vary from school to school and from year to year, even within systems that have nominally standardized rules.

Unlike other high-stakes exams such as the ACT and SAT, no significant test-preparation market has developed around college placement exams, even though hundreds of thousands of students take them each year. The reason is that many students

⁴ For comparison, the SAT takes 3 hours and 20 minutes (excluding experimental sections) and the ACT takes 2 hours and 55 minutes.

are not even aware of these exams and their consequences until after admission. A recent study that included student focus groups, counselor interviews, and a survey of matriculation officers in California concluded that students are generally uninformed about placement assessments (Venezia, Bracco, & Nodine, 2010). The study found that test preparation resources varied from college to college, that staff sometimes downplayed the consequences of the exams, and that some students even thought it would be “cheating” to prepare. The authors quote one student who reported, “[The woman at the test center] said, ‘It doesn’t matter how you place. It’s just to see where you are.’ Looking back, that’s not true. It’s really important” (Venezia et al., 2010, p. 10).

3. Theoretical Background and Previous Research

3.1 Concepts of Test Validity

In the most recent edition of the *Standards for Educational and Psychological Testing*, published by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), test validity is defined as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. ... It is the interpretation of test scores *required by proposed uses* [emphasis added] that are evaluated, not the test itself” (as cited in Brennan, 2006, p. 2). Similarly, Kane (2006) states, “It is not the test that is validated and it is not the test scores that are validated. It is the claims and decisions based on the test results that are validated” (pp. 59–60). This reflects the emphasis in modern validation theory on arguments, decisions, and consequences rather than the mere correspondence of test scores to outcomes (criteria) of interest and is what Kane (1992) calls an “argument-based approach” to validity.

The reference manuals for both major tests follow this approach and identify some of the key assumptions underpinning the validity argument for the use of test scores for course placement. For example, both the COMPASS and ACCUPLACER manuals explain that to be valid, their tests must (1) actually measure what they purport to measure, (2) they must reliably distinguish between students likely or not likely to do

well in specific “target” courses, and (3) there should be a positive statistical relationship between test scores and grades in the target courses (ACT, Inc., 2006, p. 100; College Board, 2003, p. A-62). The latter two elements relate to predictive validity, which is the focus of the current analysis.

Both manuals are explicit, however, that while predictive validity is *necessary* to demonstrate the overall validity of a test, it is not *sufficient*. As the ACCUPLACER manual warns, “Ultimately, it is the responsibility of the users of a test to evaluate this evidence to ensure the test is appropriate for the purpose(s) for which it is being used” (College Board, 2003, p. A-62). What else is required to demonstrate the valid use of a test for a given purpose? Sawyer and Schiel (2000) of ACT, Inc., argue that one must show not only that test scores are predictive of success along the desired dimension but also that “the remedial course is effective in teaching students the required knowledge and skills” (p. 4). In other words: Do students with low scores actually benefit from being assigned to remediation on the basis of this test? Simply confirming that a placement exam predicts performance in college-level math does not, on its own, imply that students with low scores should be assigned to remedial math.

Thus, even if an exam has high predictive validity, evaluations of the impact of remediation (or other support services provided on the basis of test scores) are ultimately needed to determine the overall validity of a placement testing system. As mentioned above, the available evidence is mixed regarding the impact of remediation, with some studies even finding evidence of negative effects at least for students near the placement cutoffs (for a review of the literature, see Hughes & Scott-Clayton, 2011). But if the exams themselves have limited predictive validity, their current use may not be justified regardless of the impact of remediation.

3.2 Evidence Regarding Predictive Validity

The traditional method of measuring predictive validity relies on correlation coefficients, where a coefficient of zero indicates no relationship between the test and the relevant outcome and a coefficient of one indicates perfect predictive power. The College Board publishes correlations coefficients relating each of the ACCUPLACER modules to measures of success in the relevant college credit-bearing course. The few published

studies of placement exam predictive validity by independent researchers have also typically relied on correlation coefficients, including Armstrong's (2000) study of an unnamed placement exam in use at three community colleges in California and Klein and Orlando's (2000) study of the City University of New York's since-abandoned Freshman Skills Assessment Test.

But correlation coefficients can be insufficiently informative or, even worse, misleading. Correlations between math test scores and grades in college-level math can be computed only for those students who place directly into college-level math. For those placed into remediation, this intervening intervention may confound the relationship between scores and future performance. Even if—or indeed, especially if—the test identifies the students most likely to succeed, this restriction of the range of variation may decrease the correlation coefficients (ACT, Inc., 2006, p. 101). Imagine, for example, the perfect test: Everyone scoring above a certain cutoff would have a 100 percent chance of success in the college course, and everyone below would have zero chance. If we look only at the outcomes of those initially placed into college-level, the correlation between scores and outcomes would be zero. In addition, computation of correlation coefficients requires other statistical assumptions that may be questionable (namely, that the relationships between scores and outcomes are linear and that errors are normally distributed; see ACT, Inc., 2006, p. 101). Even aside from these concerns, there is no obvious or absolute standard for how large a correlation coefficient should be to be considered sufficiently predictive.

In an effort to provide more useful measures, both the College Board and ACT, Inc., compute “placement accuracy rates,” as advocated by Sawyer (1996). This procedure starts by acknowledging that no placement rule can avoid making some mistakes—some students who could have succeeded in the college-level course will be placed into remediation (an underplacement, or Type II error), while some students who cannot succeed at the college level will be placed there anyway (an overplacement, or Type I error). Placement accuracy rates combine data on overplacements (which can be directly observed from course outcomes) and underplacements (which must be predicted from the data) to estimate what percentage of students are predicted to be accurately

placed—whether into remediation or college-level courses—under a given placement rule and definition of success.

The first step in computing these rates is to define a measure of success, such as earning a grade of B or higher in college-level math. Next, logistic regression is used to estimate the relationship between test scores and the probability of success for those students who score high enough to place into the college-level course. Third, this relationship is extrapolated to students scoring below the cutoff. Finally, for different placement rules (which may involve only a test score or may involve multiple measures), the placement accuracy rate is calculated as the sum of “observed true positives”—students who are placed at the college level and actually succeed there—and “predicted true negatives”—students who are not predicted to succeed at the college level and are “correctly” placed into remediation.

A summary of the evidence on the predictive validity of the two major placement exams is provided by Hughes and Scott-Clayton (2011). Observed correlation coefficients (available only for the ACCUPLACER) are generally higher for the math exams than for reading/writing, and are generally higher for a B-or-higher success criterion than for a C-or-higher criterion. Placement accuracy rates (available for both the COMPASS and the ACCUPLACER) generally range between 60 percent and 80 percent and show less of a pattern across test types and outcome criteria.

In addition to placement accuracy rates, ACT, Inc. (2006) also estimates the incremental validity of the COMPASS, or the typical increase in accuracy rates above what would result if all students were assigned to the college-level course. Interestingly, results indicate substantial increases in accuracy rates under the B-or-higher criterion but generally small increases in accuracy rates under the C-or-higher criterion; except for placement into college algebra, using the test with the C-or-higher criterion increased placement accuracy by only 2–6 percentage points.

It would also be useful to consider the incremental validity of test scores compared to other potential measures of college readiness, though the test makers do not provide such analyses. According to a review by Noble et al. (2004), “Using multiple measures to determine students’ preparedness for college significantly increases placement accuracy. ... For example, test scores and high school grades may be used

jointly to identify students who are ready for college-level work” (p. 302). The incremental validity of placement exams, instead of or in addition to other measures of prior achievement, is something I explore in the empirical analysis below.

A limitation of placement accuracy computations is that they require extrapolation of the relationship between test scores and outcomes (observed only for those placing directly into college-level) to those scoring below the cutoff. It thus matters whether 25 percent score above the cutoff or 75 percent do. If a relatively small proportion of students place directly into college-level, this decreases the precision of the resulting placement accuracy rates (Sawyer, 1996). There is no way to be sure that the observed relationship between scores and outcomes for high-scorers is equally applicable to very low-scorers. Sawyer (1996) concludes that as long as “25 percent or fewer of the students are assigned to the remedial course, then the [placement accuracy] procedure described here will estimate the conditional probability of success with reasonable accuracy” (p. 280), but this standard does not appear to have been met in most cases. In the ACT, Inc. (2006) study, the percentage assigned to the lower-level course was never lower than 46 percent. In many cases (including at LUCCS, as will be shown below), the proportion is much higher.

One way to address this concern is to limit the scope of the placement accuracy analysis by excluding students who score far below the cutoff. For these students, predicted rates of success in the college-level course are highly speculative. And realistically, policymakers are unlikely to consider placing these very low scorers into college-level coursework under any scenario. While an analysis excluding very low scorers may be more limited in its conclusions, it may also be more relevant for policy. I will present results from such an approach as a sensitivity analysis below.

4. The Institutional Context and Data

4.1 Institutional Context

For the period under study in this report, LUCCS was using the COMPASS numerical skills/pre-algebra, algebra, and reading exams for remedial placement, as well

as a writing exam that LUCCS adapted slightly from the standard COMPASS writing module and that LUCCS grades in-house. The two math exams are taken together, and the reading/writing exams are taken together.

The LUCCS central office establishes minimum cut scores for access to college-level courses that apply to all of the LUCCS institutions; however, schools are free to establish higher cutoffs, and some schools in some years were allowed to have lower cutoffs on the writing exam on a pilot basis. As in many systems, students are exempt from the placement exams if they score above a certain level either on the SAT, ACT, or on a standardized state high school exam; all other students must take the exams prior to the first semester of enrollment. The retesting policy is strict: students may not retake a placement exam until they have completed either a remedial course or at least 20 hours of documented participation in an alternative intervention, which might include a workshop or regular tutoring.

Students are encouraged to begin their remedial coursework right away. Although they may be able to access some college-level courses before completing remediation, students must pass college-level freshman composition and at least one credit-bearing math course in order to earn any degree, so a student cannot graduate without successfully exiting remediation. During the period under study, students needed both to pass the remedial course and retake and pass the relevant COMPASS exam in order to exit remediation.

Students' compliance with course placement decisions appears to be higher at LUCCS institutions than at many others, including those with nominally "mandatory" placement (see Bailey et al., 2010, for estimates of the rate of compliance with placement recommendations). While some students may not enroll in the required remedial course immediately, relatively few students who are assigned to a remedial course circumvent that placement to enroll in a college-level course.

4.2 Data and Sample

The data for this analysis were provided under a restricted-use agreement with LUCCS. This analysis will focus on four cohorts of first-time degree-seekers,

representing nearly 70,000 students, who entered one of LUCCS's community colleges between the fall of 2004 and fall of 2007.

Table 1 provides demographic information on the full sample and main subsamples for the predictive validity analysis. The first column describes the overall population of first-time degree-seeking entrants to LUCCS between fall 2004 and fall 2007. The second column is limited to the 80 percent of these students who took a placement exam in math (that is, excluding those who were exempt because of their scores on ACT, SAT, or standardized high school exams). The third column is limited further to those students who took the math placement exams and had information on high school math courses and grades available. Note that these students tend to be younger and are more likely to have entered college directly from high school.⁵ The fourth column is limited to the 75 percent of all entrants who took a placement exam in reading or writing, and the final column further limits this group to those that had information on high school English college preparatory courses and grades.

Like at higher education institutions generally, nearly six out of 10 LUCCS entrants are female. While more than half of LUCCS entrants are age 19 or under and come directly from high school, nearly one-quarter are 22 or older, and on average entrants are 2.6 years out of high school. Finally, LUCCS is highly diverse: over a third of students are Hispanic, over a quarter are Black (non-Hispanic), 10 percent are Asian, and 8 percent identify as other non-Caucasian ethnicities. Only 14 percent of students are White. A full 44 percent are identified (either via self-report or via a writing placement exam) as non-native English speakers.⁶

⁵ High school background information, such as grades and college-preparatory units completed in each subject, are collected by LUCCS for students who apply through a centralized application process. Students who simply show up on a given campus are known as "direct admits" and typically have much more limited background information available in the system-wide database.

⁶ Unfortunately, self-reported language status is missing for approximately one-third of the sample, and it is possible that native English speakers are more likely to have missing data on this question. Thus, we create a combined measure that identifies a student as non-native English speaking if they were flagged as such on a writing placement exam *or* if they self-reported this status on their application. Approximately 25% are flagged as ESL students after taking the writing exam, while approximately one-third self-report as non-native English speakers (or 50% of those who answered the question).

Table 1
LUCCS Degree-Seeking Two-Year Entrants: Selected Student Demographics by Data Subgroup

	LUCCS Overall	Subgroup with Math Test Score Data	Subgroup with Math Test Score and HS Math Data	Subgroup with Reading and Writing Test Score Data	Subgroup with Reading and Writing Test Score and HS English Data
Gender					
Female	0.568	0.573	0.582	0.567	0.572
Age					
Average age	21.0	21.1	20.8	21.5	21.2
18 or less	0.421	0.400	0.439	0.362	0.395
19	0.185	0.187	0.181	0.187	0.183
20	0.101	0.104	0.097	0.110	0.104
21	0.059	0.062	0.056	0.067	0.063
22 or more	0.234	0.247	0.227	0.275	0.256
Race/ethnicity					
White	0.139	0.139	0.148	0.126	0.134
Black	0.284	0.299	0.288	0.293	0.281
Hispanic	0.345	0.341	0.342	0.350	0.350
Asian	0.109	0.100	0.104	0.113	0.120
Other	0.079	0.078	0.073	0.077	0.072
Unknown	0.044	0.044	0.045	0.041	0.043
Time to college enrollment					
Years since high school graduation	2.614	2.714	2.176	2.966	2.404
Entered less than 1 year after high school graduation	0.550	0.534	0.628	0.501	0.593
Language background					
Non-native English speaker	0.485	0.475	0.477	0.515	0.517
Flagged on any pretest as ESL	0.252	0.259	0.259	0.330	0.333
Any indication of ESL or non-native speaker	0.456	0.454	0.522	0.505	0.572
Assignment to developmental education					
Math	0.630	0.789	0.778	0.701	0.685
Writing (including ESL)	0.554	0.593	0.574	0.722	0.713
Reading	0.216	0.242	0.231	0.276	0.271
Any subject	0.758	0.879	0.870	0.854	0.846
Sample size	68,220	54,412	37,860	50,576	34,808

The bottom of Table 1 indicates the percentage of each of these samples who were assigned to remedial coursework in each subject as a result of their placement exam scores.⁷ Across these four cohorts of entrants, more than three-quarters were assigned to remediation in at least one subject: 63 percent in math, 55 percent in writing, and 22 percent in reading. The proportions among those who actually take the placement exams is necessarily higher, with 78 percent of math test takers assigned to math remediation, 72 percent of reading/writing test takers assigned to writing remediation, and 28 percent of reading/writing test takers assigned to reading remediation. These high proportions of students assigned to remediation present a challenge for any analysis of predictive validity, which necessarily must rely heavily upon the patterns observed among students who place directly into college-level coursework.

5. Empirical Strategy and Main Results

5.1 Predictor Variables and Success Criteria

The previous literature on placement assessment, including the reference manuals of the test makers themselves, has emphasized the potential importance of multiple measures of readiness (College Board, 2003, p. A-2; Noble et al., 2004). However, for non-exempt students, few schools nationally appear to use multiple measures in a systematic way, perhaps because of uncertainty regarding how to collect this information efficiently or how to combine it into a simple and scalable placement algorithm (Hughes & Scott-Clayton, 2011).⁸ Thus an important goal of this study is not only to evaluate the predictive validity of the test scores currently used by LUCCS to make placement decisions, but to compare this with the predictive value of other measures that could be used either instead of or in addition to placement scores.

⁷ The percentages in this table reflect the actual assignments based on local (school-level) placement rules for the relevant entry cohort, which may be different from central LUCCS policy.

⁸ At most institutions, students who score highly enough on the ACT or SAT are exempted from the remedial placement process. LUCCS additionally exempts those who score highly enough on standardized high school exams in English and math. These exemption rules can themselves be considered a form of multiple measures, which will be examined in future work.

I focus on four alternative sets of predictor variables:

- 1) scores on the relevant placement exams (numerical skills/pre-algebra and algebra scores for math placement; reading and writing scores for English placement);
- 2) high school cumulative grade point averages, both overall and in the relevant subject; cumulative numbers of college-preparatory units completed, both overall and in the relevant subject; and indicators of whether any college-preparatory units were completed, both overall and in the relevant subject;
- 3) a combination of both (1) and (2); and
- 4) a combination of (1) and (2) plus two additional demographic predictors, whether the student graduated from a non-local high school and the number of years since high school graduation.

The two demographic predictors in variable set (4) are included as gross (but easily measurable) proxies of student motivation and maturity. Students who are returning to college after several years away from school, or who are seeking to enroll after migrating to the metropolitan area, may have higher levels of motivation and maturity on average than local students who just graduated from high school, for whom LUCCS enrollment may be more of a default next step than an active decision. I do not consider demographic variables such as gender, age, race, or ethnicity, which may have predictive value but would be unethical to consider in placement decisions.

I focus on three primary success criteria:

- 1) whether the student earns a B or better in the first college-level course taken in the relevant subject,
- 2) whether the student earns a C or better in the first college-level course taken in the relevant subject, and
- 3) whether the student passes the first college-level course taken (at LUCCS, this requires earning a D- or better) in the relevant subject.

For some analyses I also examine a continuous measure of grades earned in the first college-level course. For all of these criteria, students who withdraw from or receive an incomplete in the college-level course are treated equivalently to students who fail. Previous studies sometimes exclude these students completely; I choose to include them because they represent a significant proportion of the sample (roughly 16 percent withdraw from their first college-level course in our sample) and because withdrawal decisions are not likely to be random, but rather may be closely linked to expectations regarding course performance.

5.2 Analysis of Variation

Despite the limitations of correlation coefficients, I compute them for two reasons: first, to enable comparison with previous research, and second, to enable comparisons across alternative sets of predictors within the sample. Even if the *levels* of the correlation coefficients are biased downward because of range restriction, it may still be reasonable to compare correlation coefficients for different sets of predictors and different success criteria that are all subject to the same range restriction.

To compute the correlation coefficients, I restrict the sample to those students who have placement exam data, who ever enrolled in a college-level course in the relevant subject (math or English), and who did not take a remedial course in that subject first. I will refer to this as the math or English “estimation sample.” I then run linear probability (OLS) models of the form:

$$(1) P(success) = \alpha + \beta_1(predictor_1) + \dots + \beta_n(predictor_n) + \varepsilon$$

I then examine value of the resulting R-squared statistic, which ranges from zero to one and indicates the proportion of variation in the success criterion that can be explained by the given set of predictor variables. The correlation coefficient is simply the square root of this statistic. Because R-squared values have a more intuitive interpretation, I present them along with the correlation coefficients. Because the primary goal of this analysis is comparative, I perform no statistical corrections for restriction of range and thus the absolute levels of these correlations should be interpreted cautiously. The results are presented in Table 2.

Table 2
Relationship of College-Level Outcomes to Alternative Sets of Predictor Variables

Sample restricted to students with high school background data				
	Placement Test Scores Only	High School GPA/Units Only	Placement Test Scores PLUS HS GPA/Units	Test Scores, HS GPA/Units, PLUS Local HS, Years Since HS
Panel A. R-Squared Statistics (Proportion of Variation Explained)				
Math				
Earned B or higher in CL ^a	0.121	0.102	0.165	0.183
Earned C or higher in CL	0.069	0.077	0.109	0.121
Passed CL (D- or higher)	0.040	0.058	0.074	0.078
Grades in first CL ^b	0.129	0.119	0.183	0.204
English				
Earned B or higher in CL	0.021	0.043	0.060	0.093
Earned C or higher in CL	0.008	0.038	0.045	0.059
Passed CL (D- or higher)	0.004	0.034	0.038	0.047
Grades in first CL	0.017	0.055	0.069	0.098
Panel B. Correlation Coefficients				
Math				
Earned B or higher in CL	0.349	0.320	0.406	0.428
Earned C or higher in CL	0.263	0.278	0.330	0.348
Passed CL (D- or higher)	0.199	0.241	0.272	0.279
Grades in first CL	0.359	0.345	0.428	0.452
English				
Earned B or higher in CL	0.147	0.207	0.244	0.305
Earned C or higher in CL	0.092	0.195	0.212	0.244
Passed CL (D- or higher)	0.064	0.185	0.194	0.216
Grades in first CL	0.132	0.234	0.262	0.313
Math sample size	6,100	6,100	6,100	6,098
English sample size	9,628	9,628	9,628	9,621

Note. Math estimation sample represents 8,211 entrants from the 2004–2007 entry cohorts who took both math placement exams and who took a gatekeeper math course without taking developmental math. English estimation sample represents 14,030 entrants from the 2004–2007 entry cohorts who took both reading and writing placement exams and who took a gatekeeper English course without taking developmental reading or writing. See text for details on predictor variable sets. Adapted from author's calculations using administrative data on first-time entrants at LUCCS institutions, fall 2004 through fall 2007.

^aCL is an abbreviation for the first college-level course.

^bGrades are on a 14-point scale where 1 is fail/withdraw and 14 is A+.

Several interesting patterns are revealed by these data. First, focusing on the first or second columns, which examine the predictive value of placement scores alone for slightly different samples, one can see that exam scores are much better predictors of math outcomes than English outcomes. The overall proportion of variation explained is 13 percent for a continuous measure of math grades, compared with only 2 percent for a continuous measure of English grades. This is consistent with the findings from previous research. Second, in both math and English and regardless of the set of predictor variables, it is easier to predict success on the B-or-higher criterion than on the C-or-higher or passed-college-level criteria. In other words, it is easier to distinguish between those likely to do very well and everyone else than it is to distinguish between those likely to do very poorly and everyone else. This is also consistent with previous research.

Third, comparing across sets of predictor variables, high school achievement measures (including grades and college preparatory units taken overall, and within the relevant subject area) alone do better than placement scores alone with the sole exception of the B-or-higher criterion in math, for which placement test scores do slightly better. This is especially true for English course outcomes; English comprehension and writing skills may simply be more difficult than math skills to measure in a brief placement exam. The advantage of using high school achievement measures is especially apparent for lower standards of success. This may be because they capture non-cognitive factors such as motivation and academic engagement that are particularly important in the lower tail of the grade distribution.

Finally, for all success criteria, combining placement exam scores and high school achievement measures improves the proportion of variation explained, often by substantial amounts. The overall proportion of variation explained is 18 percent for a continuous measure of math grades, compared with 13 percent using placement exam scores alone and 12 percent using high school background alone. This increases to 20 percent with the addition of two demographic variables (an indicator for local high school and number of years since high school). In English, combining exam scores and high school achievement explains 7 percent of variation in grades, compared with 2

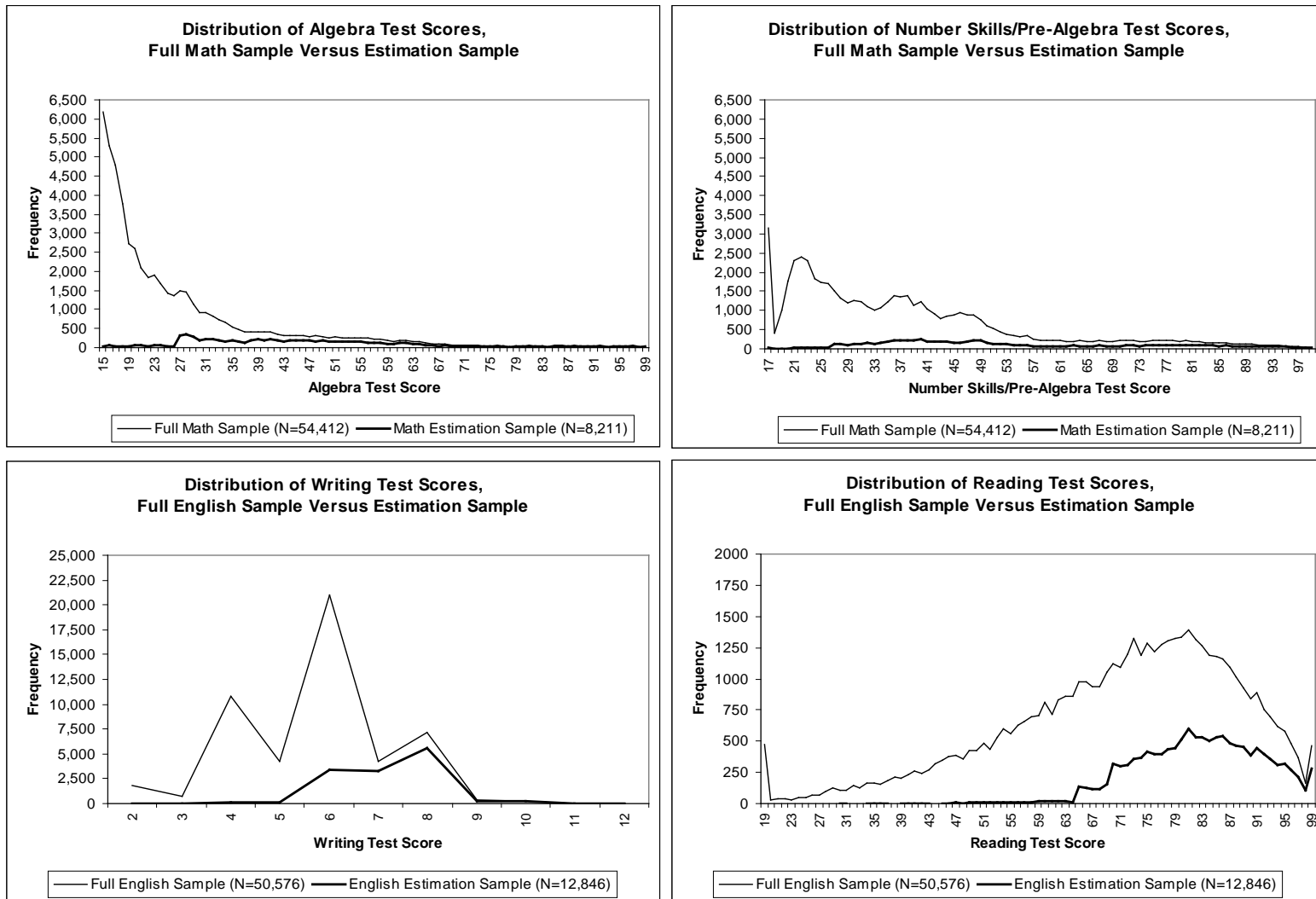
percent using scores alone and 6 percent using high school measures alone. This increases to 10 percent when additional demographic measures are included.

A final conclusion one might be tempted to draw from Table 2 is that, in absolute terms, the predictive validity of placement exam scores alone is low. Again, however, it is difficult to interpret the absolute levels given the restricted range over which these statistics must be computed. Figure 1 illustrates the restriction of range problem by showing the distribution of algebra and writing scores for the full math/English test taker samples and for the corresponding estimation samples. The overall distribution of algebra test scores (the more difficult of the two math exams) is strongly skewed, with 42 percent scoring below a 20 and 66 percent scoring below 27 (the lowest cutoff for remediation during this analysis period). Out of all math test takers, only 8,211 (15 percent) took a college-level math course without taking a developmental math course first, and of these, 90 percent had algebra test scores of 27 or higher.⁹

The range restriction is not as bad in English, where a higher proportion of students pass the placement exams, a higher proportion of students below the official cutoff are allowed directly into the college-level course, and finally, where students are more likely to actually take the college-level course than in math, conditional on eligibility. Out of all reading and writing test takers, 76 percent scored below a 7 on the writing exam (the more difficult of the two exams) but only 35 percent scored below a 6. Of all reading/writing test takers, 25 percent took a college-level English course without taking developmental reading or writing first, and of these, 72 percent had writing scores of 7 or higher.

⁹ A small number of students with lower test scores are able to take college-level courses because they qualified for an exemption based on their ACT, SAT, or standardized high school exam scores, but took a placement test anyway. Many students who score above the placement cutoffs will not be in the math estimation sample because (1) they never enrolled in the college-level course, even though they were eligible, (2) they were assigned to developmental math courses because of higher local cutoffs, or (3) they failed the pre-algebra exam (note that Figure 1 only displays the distribution of algebra scores, the more difficult of the two math modules).

Figure 1
Distribution of Test Scores for All Test-Takers Versus Those in Estimation Samples



5.3 Predicted Placement Accuracy Rates and Incremental Accuracy Rates of Placement Tests

While the analysis of variation provides some preliminary indications of the validity of placement exams in comparison with other potential sets of predictor variables, placement accuracy rates may be more useful. They do not depend on linearity or normality assumptions, they provide estimates of the proportion of students likely to succeed under different placement rules, and they enable policymakers to incorporate information regarding the costs of different types of placement “mistakes.” (They do not, however, solve the fundamental problem of range restriction—they still rely on extrapolations of a relationship observed at and above the test score cutoff to students who may have scored at the extreme low end on the test. In a sensitivity analysis presented below, I also calculate placement accuracy rates only for a sample of students scoring just above or just below the score cutoff, in which extrapolation is less of a concern.)

To compute placement accuracy rates, I again begin with an estimation sample: those students who have complete data (including test scores, high school background information, and demographic information), who ever enrolled in a college-level course in the relevant subject (math or English), and who did not take a remedial course in that subject first. I then run regressions similar to equation (1) above, but using a non-linear probit model instead of ordinary least squares (OLS). Using the parameters estimated by the probit model, I calculate predicted probabilities of success in the college-level course for the full sample of test takers. To obtain the best possible prediction, I include the full set of predictor variables (set [4] described above) and augment the model further with demographic variables (age, race/ethnicity, and a flag for whether the student was a non-native English speaker). Even though some of these variables cannot be used in a placement algorithm, they can still be used to estimate the accuracy of a more restricted placement algorithm.

Students can then be categorized into four groups, as indicated in Figure 2. Depending upon their actual placement and their predicted probability of success in the college-level course, students are either underplaced, overplaced, or accurately placed in

either the remedial or college-level course.¹⁰ The overall placement accuracy rate can then be calculated as the percentage of students placed into developmental and not predicted to succeed at college-level, plus the percentage of students placed into college-level and predicted to succeed there—that is, the sum of cells (2) and (3) in Figure 2.

Figure 2
Categorizations Based on Predicted Outcomes and Placement Decisions

Placement Decision	Predicted to Succeed in College-Level Course?	
	Yes	No
Placed into developmental ed.	(1) false negative Type II error (underplaced)	(2) accurately placed
Placed into college-level	(3) accurately placed	(4) false positive Type I error (overplaced)

Predicted success rates can also be plotted against placement exam scores to get a visual representation of the strength of their relationship, with steeper lines indicating a stronger relationship. Figure 3 plots the predicted probability of success in college-level math against math exam scores, under three alternative success criteria. The LUCCS minimum cutoff (in place at the end of the analysis period) is indicated by the vertical line at 30. Predicted rates of success are obviously lowest using the B-or-higher criterion, but the slope of the line is also steepest for this criterion, consistent with the pattern of correlation coefficients found above. Figure 4 does the same for college-level English, with similar patterns evident.

¹⁰ Previous research has considered students as “likely to succeed” if the estimated probability of success generated by the non-linear regression is at least 50% (see, e.g., Mattern & Packman, 2009). However, because this information is ultimately aggregated to the group level, there is no need to explicitly assign each student to a single cell. Instead, one can simply take the average of these individual predicted probabilities to estimate predicted *rates* of success in the college-level course for (1) those placed into developmental education and (2) those placed directly into college-level.

Figure 3
Probability of Gatekeeper Success, by Math Part 2 Scores

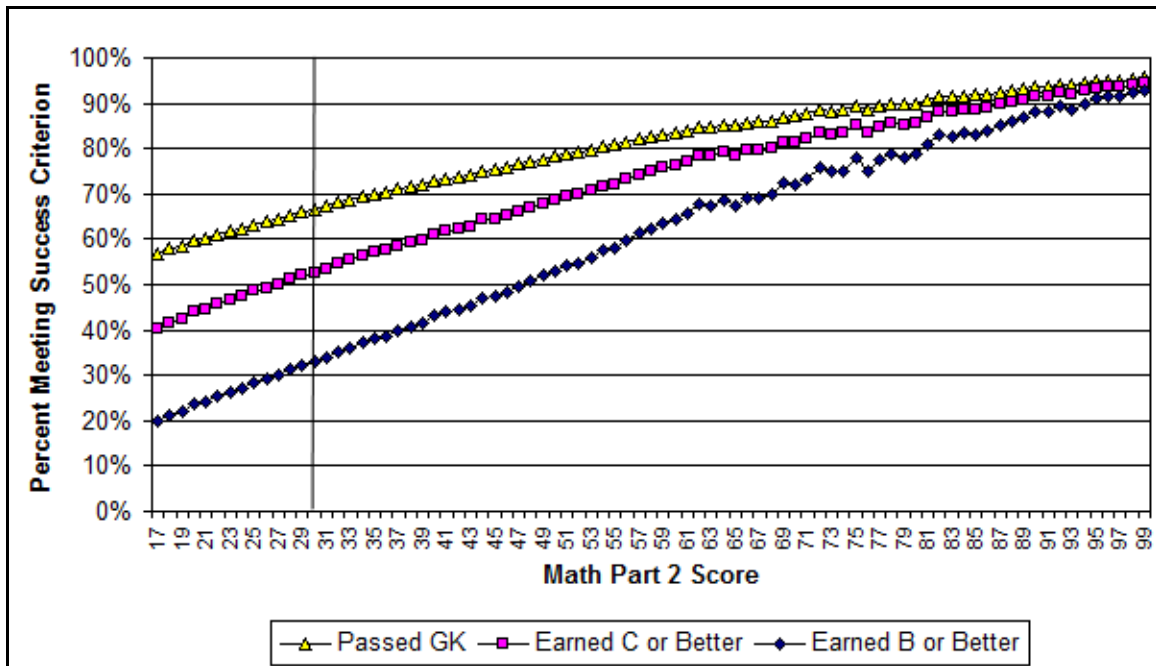
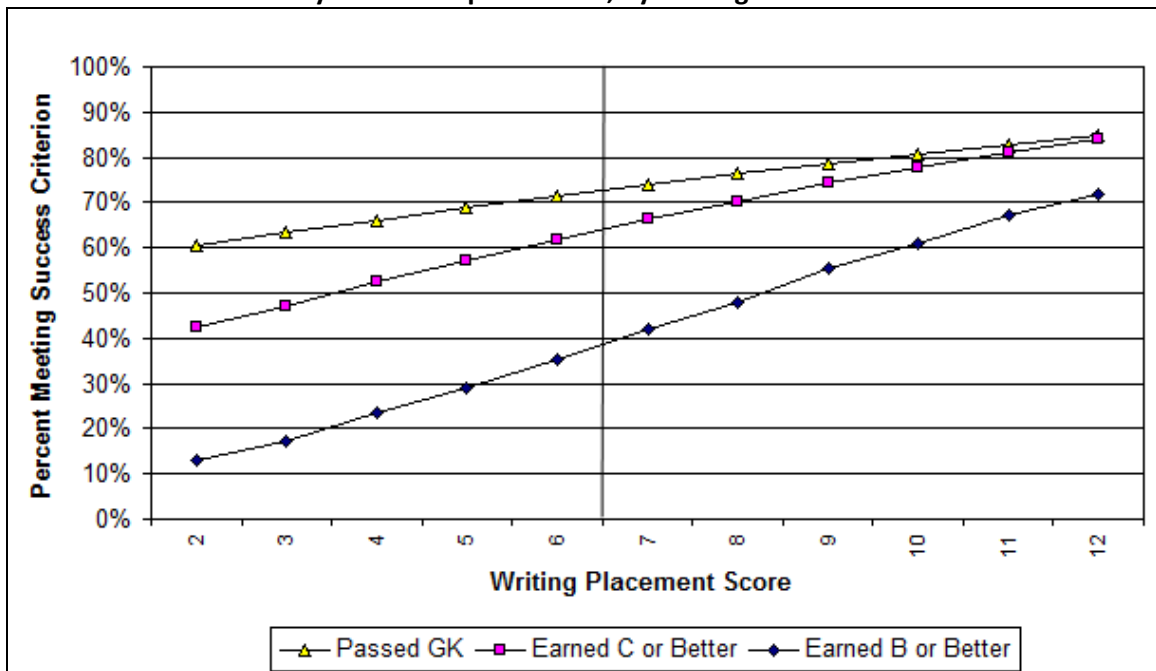


Figure 4
Probability of Gatekeeper Success, by Writing Placement Scores



An interesting feature of these graphs is that they can be used to determine “optimal” placement score cutoffs, depending upon policymakers’ chosen success criterion and their relative valuations of the costs of Type I (overplacement) and Type II (underplacement) errors. If policymakers weight overplacement and underplacement errors equally, then the optimal cutoff occurs at the score where the probability of college-level success is 50 percent. If the probability of success is higher at the chosen cutoff, then students just above the cutoff have a higher probability of being accurately placed than those just below the cutoff, so moving the cutoff down would increase overall accuracy. This would imply an optimal cutoff of approximately 47 on the algebra placement exam for the B-or-higher criterion or 26 for the C-or-higher criterion. For the passing criterion, placement accuracy would be maximized by allowing all students to take the college-level course.

The overall predicted accuracy rates using the LUCCS cutoffs in place at the end of the analysis period are computed in the first column of Table 3. The next two columns indicate the predicted accuracy rates that would result under two hypothetical (and extreme) alternative placement policies, if test scores were not available: either placing all students into developmental or all students into college-level math. The bolded numbers indicate the policy that results in the highest overall accuracy for a given success criterion. The final two columns indicate the incremental accuracy of using placement tests instead of nothing at all, with the numbers in bold representing the incremental accuracy of placement tests versus the next best alternative (placing all students into either developmental or all into college-level).

Table 3
Predicted Placement Accuracy Rates Using Placement Test Scores,
Versus Placing All Students in College Level or Remedial

	Accuracy Rate, Using Placement Test Cutoffs	Accuracy Rate, All Students In Developmental	Accuracy Rate, All Students In College Level	Incremental Validity vs. All Dev Ed	Incremental Validity vs. All Coll. Lev
Math					
Earned B or higher in GK	0.695	0.695	0.305	0.000	0.390
Earned C or higher in GK	0.582	0.505	0.495	0.077	0.087
Passed GK (D- or higher)	0.493	0.361	0.639	0.131	-0.146
English					
Earned B or higher in GK	0.613	0.661	0.339	-0.048	0.274
Earned C or higher in GK	0.433	0.395	0.605	0.038	-0.172
Passed GK (D- or higher)	0.361	0.294	0.706	0.067	-0.345

Note. Math estimation sample includes 6,100 entrants from the 2004–2007 entry cohorts who took a gatekeeper math course without taking developmental math, and who have placement test scores and high school background data available. Math prediction sample includes all 37,860 entrants from 2004–2007 who have both placement test scores and high school background information. English estimation sample includes 9,628 entrants from the 2004–2007 entry cohorts who took a gatekeeper English course without taking developmental English/reading, and who have placement test scores and high school background data available. English prediction sample includes all 36,917 entrants from 2004–2007 who have both placement test scores and high school background information. Placement accuracy rates are calculated as the percentage of students who are predicted to succeed in the gatekeeper class and are accurately placed there, or are predicted not to succeed in the gatekeeper course and are accurately placed in developmental education. Adapted from author’s calculations using administrative data on first-time entrants at LUCCS institutions, fall 2004 through fall 2007.

Though placement accuracy rates are meant to be more transparent than the correlation coefficient, the results tell a somewhat confusing story. First, focusing just on the first column of Table 3, accuracy rates are better for the higher success criteria and are higher in math than in English, consistent with the patterns found above. But looking at the accuracy rates in the next two columns indicates that in most cases, similar or even higher accuracy rates could have been achieved without using the placement exams at all, but instead by assigning all students to either the developmental or college-level course. The greatest gain in incremental accuracy occurs for the C-or-higher criterion in math, for which using the placement test cutoffs increases accuracy by 8 percentage points (or about 16 percent) compared with assigning everyone to the same level. But in several other cases, using placement exams as a screen actually results in substantially *lower* accuracy rates than using nothing at all; in other words, the increase in the number of qualified students who are prevented from accessing college-level with the exams

outweighs the decrease in the number of unqualified students who are admitted into college-level courses.

One aspect that is particularly unhelpful about these findings is that the policy conclusions depend enormously upon which particular success criteria is chosen, though in practice all three criteria may have some value. For example, if policymakers only care about the B-or-higher criterion, then using the current cutoffs or assigning all students to developmental achieve virtually identical accuracy rates. For the C-or-higher criterion, the current cutoffs are best in math while assigning all students to college-level is best in English. For the passing criterion, assigning all students to college-level is the accuracy-rate maximizing policy in both subjects.

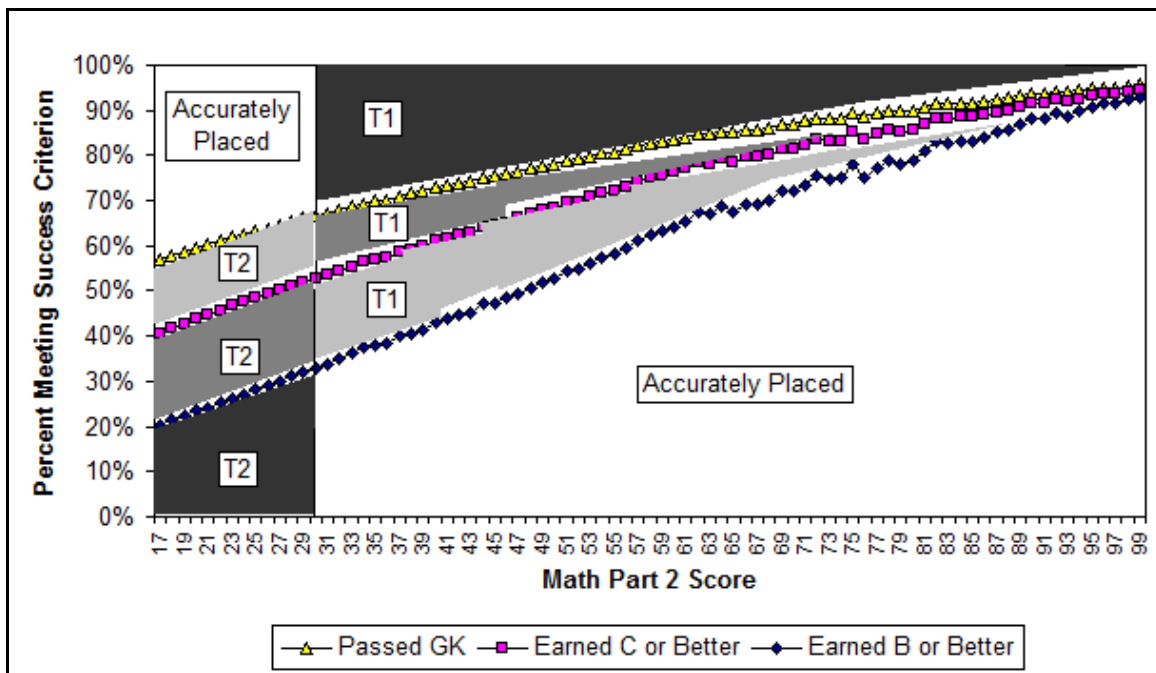
5.4 All Mistakes Are Not Equal: Minimizing the Severe Error Rate and Other Considerations

Defining multiple measures of placement validity. One way to make the analysis more useful and realistic is to recognize that all types of placement mistakes are not created equal. Under the B-or-higher criterion, for example, an underplacement (Type II) error may be much worse than an overplacement (Type I) error. In other words, we may be very concerned if many students who could have earned at least a B are wrongly placed into developmental, but less concerned if many students who are placed in college-level end up earning a C instead of a B. Conversely, under the passing criterion, we may be more concerned about overplacement versus underplacement: the cost of a student *failing* the college-level class may be much worse than “wrongly” assigning someone to developmental coursework if they would have just barely passed at the college level.

Figure 5 divides students graphically into those that are predicted to be accurately placed regardless of the success criteria and those that are predicted to be placement “mistakes” of varying severity. Type I (overplacement) and Type II (underplacement) errors are indicated with “T1” and “T2;” more severe errors are shaded in darker tones. Policymakers could assign different weights to each region in this chart and then choose the policy that minimizes the sum of severity-weighted errors, rather than focusing on the simple sum of Type I and Type II errors. The social cost of different types of errors

would include both financial and psychic costs to misplaced students, as well as the potential externalities borne by instructors and classmates of misplaced students. The weights may also reflect that estimates of overplacements are more reliable than predictions of underplacements (again, because the latter rely on statistical extrapolation). One simple weighting scheme is to focus only on the most severe errors, shaded in dark grey in Figure 5: students predicted to earn a B or better in college-level but instead placed into remediation, and students who were placed into college-level but failed there. I refer to this as the *severe error rate*.

Figure 5
Probability of Gatekeeper Success, by Math Part 2 Scores



Policymakers in practice may want to give weight to additional considerations beyond the severe error rate. For example, given two different placement systems with the same overall error rates, policymakers likely will prefer the system that assigns fewer students to remediation and that has a higher success rate in the college-level course. Rather than presuming how policymakers should weight placement accuracy rates against remediation rates and college-course pass rates, I simply compute the overall percentage of students assigned to remediation as well as the percent succeeding (using the C-or-higher criterion) among those placed directly into college-level. Finally, I compute the

overall percentage of students who are both placed directly into college-level *and* predicted to succeed there (again under the C-or-higher criterion).

Table 4 computes these additional measures of usefulness under the current placement test score cutoffs and under the alternatives of placing all students in either developmental or college-level. The final two columns compute the incremental change in each of these measures that results from using placement tests as a screen.

Table 4
Predicted Severe Error Rates Using Placement Test Scores,
Versus Placing All Students in College Level or Remedial

	(1) Using Placement Test Cutoffs	(2) All Students In Developmental	(3) All Students In College Level
Math			
Severe error rate	0.240	0.305	0.361
<i>Severe overplacement rate</i>	<i>0.058</i>	<i>0.000</i>	<i>0.361</i>
<i>Severe underplacement rate</i>	<i>0.183</i>	<i>0.305</i>	<i>0.000</i>
Remediation rate	0.748	1.000	0.000
College-level success rate (C or above), for those assigned to college level	0.670	n/a	0.495
Immediate college-level success rate, for all those taking tests ^a	0.169	0.000	0.495
English			
Severe error rate	0.334	0.339	0.294
<i>Severe overplacement rate</i>	<i>0.045</i>	<i>0.000</i>	<i>0.294</i>
<i>Severe underplacement rate</i>	<i>0.289</i>	<i>0.339</i>	<i>0.000</i>
Remediation rate	0.805	1.000	0.000
College-level success rate (C or above), for those assigned to college level	0.716	n/a	0.605
Immediate college-level success rate, for all those taking tests ^a	0.140	0.000	0.605

Note. The severe error rate is the sum of the proportion of students 1) placed into college level and predicted to fail there and 2) placed into remediation although they were predicted to earn a B in the college level. The remediation rate is the percentage of all students assigned to remediation. Adapted from author's calculations using administrative data on first-time entrants at LUCCS institutions, fall 2004 through fall 2007.

^aThe overall college-level success rate is the percentage of all students who are both assigned directly to college level and predicted to earn at least a C grade there. It does not account for students who may eventually succeed in college level after completing a remedial sequence.

The results indicate that compared with placing all students into developmental education, using placement tests significantly improves placement outcomes regardless of how these different measures are weighted (see fourth column). The severe error rate is 7 percentage points lower in math (24 percent versus 31 percent) and slightly lower in English (33 percent versus 34 percent) than what would result if all students were assigned to remediation. Obviously, significantly fewer students are assigned to remediation (25 percentage point and 19 percentage point reductions in math and English, respectively) and as a result, a higher proportion of students immediately succeed in the college-level course.

The usefulness of these placement tests is more mixed when compared against assigning all students directly to college-level coursework. In math, using the placement tests results in a substantial 12 percentage point reduction in the severe error rate, as well as an 18 percentage point increase in college-level success rates (among those placed directly into college-level). But because of the enormous 75 percentage point increase in remediation, the use of placement tests reduces the overall proportion of students *immediately* assigned to and succeeding in college-level math by 33 percentage points. While the hope is that many of these students will eventually progress through remediation and successfully complete college-level coursework later, previous research has indicated this often does not happen (Jaggars & Hodara, 2011; Bailey et al., 2010).

In English, the sole benefit of placement exams appears to be to increase the success rates in college-level coursework, among those placing directly into college-level, by 11 percentage points (from 61 percent to 72 percent). This measure may be particularly important to instructors, who may find it disruptive if too many students in their classes have very low probabilities of success. But these tests generate virtually no reduction in the overall severe error rate (in other words, while the placement tests reduce severe overplacements, they increase severe underplacements by the same amount), while at the same time dramatically increasing the proportion of students assigned to remediation and reducing the overall proportion immediately succeeding at the college-level.

Restricting the sample to students near the placement test cutoffs. A critique of the above analysis is that the underlying model predicting each student's probability of

success in the college-level course relies heavily on extrapolation from the experiences of students above the cutoff to students far below the cutoff. It may be both unrealistic and unwise to expect policymakers to consider a dramatic change in policy—such as assigning all students to college-level work—given the level of uncertainty about how students far below the cutoffs might perform. Thus, I examine these four measures of usefulness for a restricted sample of students just above and just below the LUCCS test score cutoffs (that is, students scoring between 25 and 34 on the algebra test and between 5 and 8 on the writing test). I also look at the consequences of assigning all of the students in this range to developmental (i.e., simulating a modest increase in score cutoffs) or assigning all students in this range to college-level (i.e., simulating a modest decrease in score cutoffs).

The results are presented in Table 5. For this restricted sample, the severe error rates are higher, while the remediation rates and college-level success rates (among those assigned to college-level) are lower. But the conclusions are essentially unchanged. Assigning all of these students to developmental education is never the best option. Assigning all students to college-level in math increases the severe error rate and lowers the success rate among those placed directly into college-level, but dramatically increases the percentage of students who are predicted to succeed at the college level in their first term. In English, the only drawback to allowing all of these “marginal” students to enter college-level directly is a modest decline in the college-level success rate (from 71 percent to 64 percent). The other three measures of placement outcomes show improvement.

Table 5
Predicted Severe Error Rates and Other Measures,
for Students Just Above and Just Below Placement Test Cutoffs

	Placement Test Scores Only	All Students In Dev. Ed.	All Students In College Lev.
Math (restricted to students +/- 5 points around algebra cutoff)			
Severe error rate	0.295	0.318	0.342
<i>Severe overplacement rate</i>	<i>0.093</i>	<i>0.000</i>	<i>0.342</i>
<i>Severe underplacement rate</i>	<i>0.202</i>	<i>0.318</i>	<i>0.000</i>
Remediation rate	0.703	1.000	0.000
College-level success rate (C or above), for those assigned to college level	0.556	n/a	0.517
Immediate college-level success rate, for all those taking tests ^a	0.165	0.000	0.517
English (restricted to students +/- 2 points around writing cutoff)			
Severe error rate	0.340	0.377	0.276
<i>Severe overplacement rate</i>	<i>0.058</i>	<i>0.000</i>	<i>0.276</i>
<i>Severe underplacement rate</i>	<i>0.281</i>	<i>0.377</i>	<i>0.000</i>
Remediation rate	0.750	1.000	0.000
College-level success rate (C or above), for those assigned to college level	0.709	n/a	0.635
Immediate college-level success rate, for all those taking tests ^a	0.177	0.000	0.635

Note. The severe error rate is the sum of the proportion of students 1) placed into college level and predicted to fail there and 2) placed into remediation although they were predicted to earn a B in the college level. The remediation rate is the percentage of all students assigned to remediation. Adapted from author's calculations using administrative data on first-time entrants at LUCCS institutions, fall 2004 through fall 2007.

^aThe overall CL success rate is the percentage of all students who are both assigned directly to college level and predicted to earn at least a C grade there. It does not account for students who may eventually succeed in college level after completing a remedial sequence.

The next section will compare placement accuracy rates and severe error rates for alternative sets of predictor variables.

5.5 Comparing Placement Outcomes Across Alternative Sets of Predictors

Policymakers have options beyond simply using or not using placement exams. A more interesting analysis is how much each dimension of placement outcomes might be improved by using high school background either instead of or in addition to placement exam scores.

I generated alternative placement algorithms by regressing college-level math and English grades (among only those assigned directly to college-level) on the three alternative sets of predictor variables described above in Section 5.1. I then used the parameters from these regressions to generate an index representing predicted college-level grades in the relevant subject for all students. Finally, I simulated placement cutoffs at the 75th percentile of predicted math grades and the 80th percentile of predicted English grades. This ensures that each placement algorithm generates the same proportion of students assigned to remediation as would the LUCCS test score cutoffs.

The results are shown in Table 6. The results indicate that compared with the current use of placement scores, using high school GPA/units alone without placement exam scores results in lower severe error rates, higher college-level success rates among those assigned directly to college-level, and higher rates of overall (immediate) college-level success in both math and English. The gains on these measures are particularly pronounced in English. Combining both high school background and test scores with two demographic measures—years since high school and whether the student graduated from a local high school—produces the best results for every dimension of placement effectiveness.

Table 6
Predicted Severe Error Rates and Other Measures, Using Alternative Measures for Placement

	(1)	(2)	(3)	(4)	(5)
	Placement Test Scores Only	Index of HS GPA/Units Only	Placement Test Scores PLUS HS GPA/Units	Test Scores, HS GPA/Units, PLUS Local HS, Years Since HS	Use Students' Best of Test Scores or HS Index
Math					
Severe error rate	0.240	0.227	0.213	0.208	0.217
Severe overplacement rate	0.058	0.048	0.045	0.044	0.074
Severe underplacement rate	0.183	0.179	0.168	0.164	0.143
Remediation rate	0.748	0.747	0.747	0.747	0.666
College-level success rate (C or above), for those assigned to college level	0.670	0.708	0.734	0.747	0.676
Immediate college-level success rate, for all those taking tests ^a	0.169	0.179	0.185	0.189	0.226
English					
Severe error rate	0.334	0.297	0.295	0.281	0.280
Severe overplacement rate	0.045	0.022	0.027	0.023	0.058
Severe underplacement rate	0.289	0.275	0.267	0.258	0.222
Remediation rate	0.805	0.798	0.798	0.797	0.690
College-level success rate (C or above), for those assigned to college level	0.716	0.821	0.815	0.844	0.758
Immediate college-level success rate, for all those taking tests ^a	0.140	0.166	0.165	0.171	0.235

Note. The severe error rate is the sum of the proportion of students 1) placed into college level and predicted to fail there (severely overplaced) and 2) placed into remediation although they were predicted to earn a B in the college level (severely underplaced). The remediation rate is the percentage of all students assigned to remediation. Alternative placement rules were generated by regressing college-level math and English grades (among those assigned directly to college level) on alternative sets of predictor variables, and then using the parameters from these regressions to generate predicted college-level grades for all students. Placement cutoffs were then established at the 75th percentile for math and the 80th percentile for English, to ensure that all placement algorithms would generate the same proportion assigned to remediation as the LUCCS cutoffs would. For column (5), students are placed into college-level courses if they score above the cutoff percentile (75th percentile in math, 80th percentile in English) on either the placement exams or the index based on high school grades and courses completed. Adapted from author's calculations using administrative data on first-time entrants at LUCCS institutions, fall 2004 through fall 2007.

^aThe overall CL success rate is the percentage of all students who are both assigned directly to college level and predicted to earn at least a C grade there. It does not account for students who may eventually succeed in CL after completing a remedial sequence.

The use of multiple measures can generate further improvements if we relax the restriction of keeping the remediation rate fixed. In column (5) of Table 6, I simulate the consequences of a more liberal policy which would allow students into college-level courses if they rank above the cutoff percentile (75th percentile in math, 80th percentile

in English) on either the placement exam or on an index of high school grades and courses completed.¹¹ Compared with the effects of using placement scores alone (column 1), this system would lower remediation rates by 8 percentage points in math and 12 percentage points in English—while also reducing the overall severe error rate and maintaining or even improving pass rates in the college-level course.

5.6 Summary of Empirical Results

Taken as a whole, the analyses above present a fairly consistent pattern of findings. First, placement test scores have much more predictive power in math than in English. Math scores alone explain about 13 percent of the variation in first college-level math course grades, while reading/writing scores explain less than 2 percent of the variation in first college-level English grades. Overall placement accuracy rates are higher in math than in English (58 percent versus 43 percent accurately placed under a C criterion of success), and severe error rates are lower (24 percent versus 33 percent). Compared with abandoning the exams and allowing all students direct access to college-level courses, using placement scores in math generates a substantial reduction in severe placement errors and a substantial increase in success rates among those placed directly into college-level. But in English, using placement scores actually increases the number of severe errors and generates only a modest increase in the success rate of those placed directly into college-level.

Second, placement test scores are better at predicting who is likely to do well in the college-level course than predicting who is likely to fail. For example, placement scores predict 12 percent of the variation in who gets a B or higher in the college level math course, but only 4 percent of the variation in who passes versus fails (the corresponding statistics in English are 2 percent and 0.4 percent, respectively). The use of placement test scores results in a full 70 percent of students being accurately placed in math under the B-or-higher success criterion, but only 49 percent under the passing criterion (the corresponding statistics for English are 61 percent and 36 percent, respectively).

¹¹ This index is the same index of predicted college math/English grades, based on high school grades and courses completed, used in column (2) of Table 6.

Third, the incremental validity of placement tests relative to high school background predictors of success is weak, even in math. Adding test scores to a model using high school GPA/units to predict college-level grades increases the proportion of variation explained by about 6 percentage points in math (to 18 percent from 12 percent) and less than 2 percentage points in English (to 7 percent from 5.5 percent). But even the improvement in the R-squared and associated correlation coefficient in math yields virtually no practical improvement in the severe error rate or in the success rate of students placed directly into the college-level course. In both math and English, using high school GPA/units *alone* as a placement screen results in better outcomes than using placement test scores alone (substantially so in English), and adding in placement test scores results in little additional improvement.

Fourth, simulations indicate that allowing students to test out of remediation based on the best of either their placement scores or high school achievement could substantially lower remediation rates (by 8 percentage points in math and 12 percentage points in English) without compromising success rates in college-level coursework.

Finally, while a rich predictive placement algorithm including test scores, high school background, and two proxies for student motivation could reduce severe placement errors by about 15 percent (from 24 to 21 percent in math, and from 33 to 28 percent in English), even this rich algorithm comes far from eliminating severe placement mistakes.

6. Discussion

6.1 Possible Explanations for the Limited Predictive Validity of Placement Exams

In math, one possible explanation for the limited predictive validity of placement exam scores may be a disconnect between the limited range of material tested on the exam and the material required to succeed in the typical first college-level math course (Jaggars & Hodara, 2011). ACT, Inc.'s own (2006) analysis suggests that the incremental validity of the COMPASS algebra exam is higher for predicting success in "college algebra" than "intermediate algebra." But many students meet their college-level math

requirement by taking courses that are not primarily algebra-based. For example, Introductory Statistics is a popular course at several LUCCS schools. At one school the most popular first college-level math course (for those placing directly into college-level) is described in the course catalogue as a “basic course in mathematical discovery. Students participate in the development and investigation of topics such as: number sequences, calculating devices, extrapolation, mathematical mosaics and curves, probability and topology.”

Similarly, many faculty members complained that the writing exam considered here was not a good measure of the general writing skills needed to succeed in college-level coursework.¹² In addition, while there is much less variation on paper in the first college-level English course that students take—it is typically a composition-based “Freshman English” course—there still may be considerable variation from school to school or instructor to instructor in terms of assignments required and standards for successful completion. Grades are notoriously more subjective in English than in math, which makes them more difficult to predict.

In both math and English, high school background measures may be more useful predictors of success in a wide range of settings because they capture both a wider range of cognitive skills than can be evaluated on a brief placement exam, and because they also incorporate non-cognitive factors such as student motivation. Alternatively, to the extent that grades at both the high school and college level may be influenced by social promotion norms, past grades may simply be a better predictor of who is likely to be socially promoted in the future (for better or worse).

And there are other limitations to relying on grades as a measure of success. Besides the fact that grades may vary across institutions, or across courses within schools, the focus on grades may also overlook other important outcomes, such as knowledge acquisition, performance in other courses, persistence, or even degree completion (though it is not clear that placement exams would be any more predictive of these alternative outcomes). Of course, the COMPASS and ACCUPLACER are not designed to predict these outcomes, and it would be unreasonable to expect a single exam to meet all needs. But because these placement exams are used not just for placement in

¹² Personal communication with LUCCS administrator, August 11, 2011.

math and English, but also serve as de facto college entry exams, their predictive validity for broader, longer-term measures of college success is an important topic for future research.

6.2 Limitations of the Use of High School Background Measures

An important caveat to the findings above is that only about 70 percent of LUCCS test takers have high school transcript information available. The remaining 30 percent without transcripts are, on average, four years out of high school; it may be impractical to expect to collect transcript data from them. Thus, there may be little alternative for some students to giving them some sort of placement exam. Self-reported high school background information could be elicited at registration; however, it is not clear whether self-reported grades and units completed, particularly for students many years out of high school, would have the same predictive power as the transcript data utilized here. Still, this would not seem to justify ignoring demonstrably useful information for the majority of the incoming student population.

It is also possible that the high school transcript data used here may be of higher quality than is typical for community colleges. LUCCS has developed rules for systematically coding which courses from the students' transcripts count as "college preparatory units" (which are the only courses considered here). Future research should investigate the predictive validity of high school transcript records more generally, as well as the validity of self-reported grades for those without transcripts.

6.3 The Salience of Different Types of Placement Mistakes

Compared with using nothing at all, the one measure on which placement exams generate consistent improvements is the success rate among students placed directly into the college-level course (see Table 4). Perhaps not coincidentally, this is one measure which is easily observable to both policymakers and practitioners on the ground. When a student is placed into a college-level course and fails there (an overplacement error), the fact that there has been a placement mistake is painfully obvious to all. Conversely, while we know that underplacement errors must occur in theory—and I have provided statistical estimates of their prevalence above—they are invisible to the naked eye. Among students who do well in a remedial course, it may be difficult for an instructor (or

even the student herself) to know whether they were appropriately placed or might have succeeded in the college-level course as well. In any case, when a student does well in a remedial course, this is unlikely to be perceived as a problem. The analysis above highlights the need for policymakers and practitioners to consider the prevalence and consequences of all types of placement errors—not just overplacements, but the less visible underplacements as well.

Still, because of the strong assumptions required to predict college-level outcomes for students at the extreme low end of the test distribution, it is right for policymakers to treat these estimates of underplacement cautiously. If all students were admitted directly to college-level courses, it is probable that the entire definition of “college-level” coursework would change. Nonetheless, the more conservative analysis presented in Table 5 (which includes only students within a few points of the cutoffs) suggests that lowering the cutoffs by just a few points would enable many more of these “marginal” students to pass a college-level course in their first semester. And the analysis in Table 6 demonstrates that the use of multiple measures can enable a system to reduce severe placement errors and improve college-level success rates, while keeping the remediation rate unchanged—or to reduce remediation rates without any adverse consequences.

6.4 The Impact of Remediation on Future College-Level Outcomes

Finally, as discussed in Section 3 above, evaluations of the impact of remediation (or other support services provided on the basis of test scores) are ultimately needed to determine the overall validity of a placement testing system. If remediation does not substantially improve remediated students’ probabilities of success, then this exacerbates the cost of underplacement mistakes and may lead policymakers to prefer strategies that place more students directly into college-level courses, even if the percentage succeeding there decreases as a result. If remediation is effective, then it may make sense to have higher rates of remediation in order to maintain high success rates in the college-level course. However, existing research suggests this is not the case, at least for students scoring near the remediation cutoff.

6.5 Summary and Conclusions

This paper has analyzed the predictive validity of the COMPASS, one of the most prevalent placement exams used nationally, using data on over 42,000 first-time entrants to a large urban community college system. Using both traditional correlation coefficients as well as more useful decision-theoretic measures of placement accuracy and error rates, I find that placement exams are more predictive of success in math than in English, and more predictive of who is likely to do well in college-level coursework than who is likely to fail. However, the rate of overplacement and underplacement mistakes are significant in both subjects (24 percent to 33 percent).

The predictive power of placement exams is in a sense quite impressive given how short they are (often taking about 20–30 minutes per subject/module). But overall the correlation between scores and later course outcomes is relatively weak, especially in light of the high stakes to which they are attached. Given that students ultimately succeed or fail in college-level courses for many reasons beyond just their performance on placement exams, it is questionable whether their use as the sole determinant of college access can be justified on the basis of anything other than consistency and efficiency. Allowing more students directly into college-level coursework (but perhaps offering different sections of college-level courses, some of which might include supplementary instruction or extra tutoring), could substantially increase the numbers of students who complete college-level coursework in the first semester, even if pass rates in those courses decline.

Even systems that are reluctant to relax their test score cutoffs for college-level work could do better than relying solely on test scores for remedial placement. Using high school achievement alone as a placement screen results in fewer severe placement mistakes than using test scores alone—substantially so in English—without changing the percentage of students assigned to remediation. In other words, if a school thinks roughly 25 percent of their incoming students can proceed directly to college-level work, using high school achievement rather than test scores better identifies the right 25 percent. Similarly, without changing remediation rates, combining both test scores, high school achievement, and selected background characteristics (years since high school graduation

and whether the student is coming from a local high school) could reduce severe placement errors by about 15 percent (or 3 to 5 percentage points) in each subject while simultaneously improving college-level success rates. Finally, allowing students to test into college-level work using the best of either their placement scores or an index of their high school background could markedly lower the remediation rate without compromising college-level success rates.

References

- ACT, Inc. (2006). *COMPASS/ESL reference manual*. Iowa City, IA: Author.
- Armstrong, W. B. (2000). The association among student success in courses, placement test scores, student background data, and instructor grading practices. *Community College Journal of Research and Practice*, 24(8), 681–695.
- Bailey, T., Jeong, D. W., & Cho, S. W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, 29(2), 255–270.
- Bettinger, E. P., & Long, B. T. (2009). Addressing the needs of underprepared students in higher education: Does college remediation work? *Journal of Human Resources*, 44(3), 736–771.
- Boatman, A. & Long, B. T. (2010). *Does remediation work for all students? How the effects of postsecondary remedial and developmental courses vary by level of academic preparation* (NCPR Working Paper). New York, NY: National Center for Postsecondary Research.
- Brennan, R. L. (Ed.). (2006). *Educational measurement* (4th ed.). Westport, CT: ACE/Praeger Publishers.
- Calcagno, J. C., & Long, B. T. (2008). *The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance* (NBER Working Paper No. 14194). Cambridge, MA: National Bureau of Economic Research.
- College Board. (2003). *ACCUPLACER OnLine: Technical manual*. New York, NY: College Board.
- College Board. (2007). *ACCUPLACER coordinator's guide*. New York, NY: College Board.
- Hughes, K. L., & Scott-Clayton, J. (2011). *Assessing developmental assessment in community colleges* (CCRC Working Paper No. 19, Assessment of Evidence Series). New York, NY: Columbia University, Teachers College, Community College Research Center.
- Jaggars, S. S., & Hodara, M. (2011). *The opposing forces that shape developmental education: Assessment, placement, and progression at CUNY community colleges* (CCRC Working Paper No. 36). New York, NY: Columbia University, Teachers College, Community College Research Center.

- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: ACE/Praeger Publishers.
- Klein, S. P., & Orlando, M. (2000). *CUNY's testing program: Characteristics, results, and implications for policy and research*. Santa Monica, CA: RAND.
- Martorell, P., & McFarlin, I. J. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *The Review of Economics and Statistics*, 93(2), 436–454.
- Mattern, K. D., & Packman, S. (2009). *Predictive validity of ACCUPLACER scores for course placement: A meta-analysis* (Research Report No. 2009-2). New York, NY: College Board.
- Noble, J. P., Schiel, J. L., & Sawyer, R. L. (2004). Assessment and college course placement: Matching students with appropriate instruction. In J. E. Wall & G. R. Walz (Eds.), *Measuring up: Assessment issues for teachers, counselors, and administrators* (pp. 297–311). Greensboro, NC: ERIC Counseling & Student Services Clearinghouse and the National Board of Certified Counselors.
- Parsad, B., Lewis, L., & Greene, B. (2003). *Remedial education at degree-granting postsecondary institutions in fall 2000: Statistical analysis report* (NCES 2004-101). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Primary Research Group, Inc. (2008). *Survey of assessment practices in higher education*. New York, NY: Author.
- Sawyer, R. (1996). Decision theory models for validating course placement tests. *Journal of Educational Measurement*, 33(3), 271–290.
- Sawyer, R. (2007). Indicators of usefulness of test scores. *Applied Measurement in Education*, 20(3), 255–271.
- Sawyer, R., & Schiel, J. (2000, April). *Posttesting students to assess the effectiveness of remedial instruction in college*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- Venezia, A., Bracco, K. R., & Nodine, T. (2010). *One shot deal? Students' perceptions of assessment and course placement in California's community colleges*. San Francisco, CA: WestEd.